

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

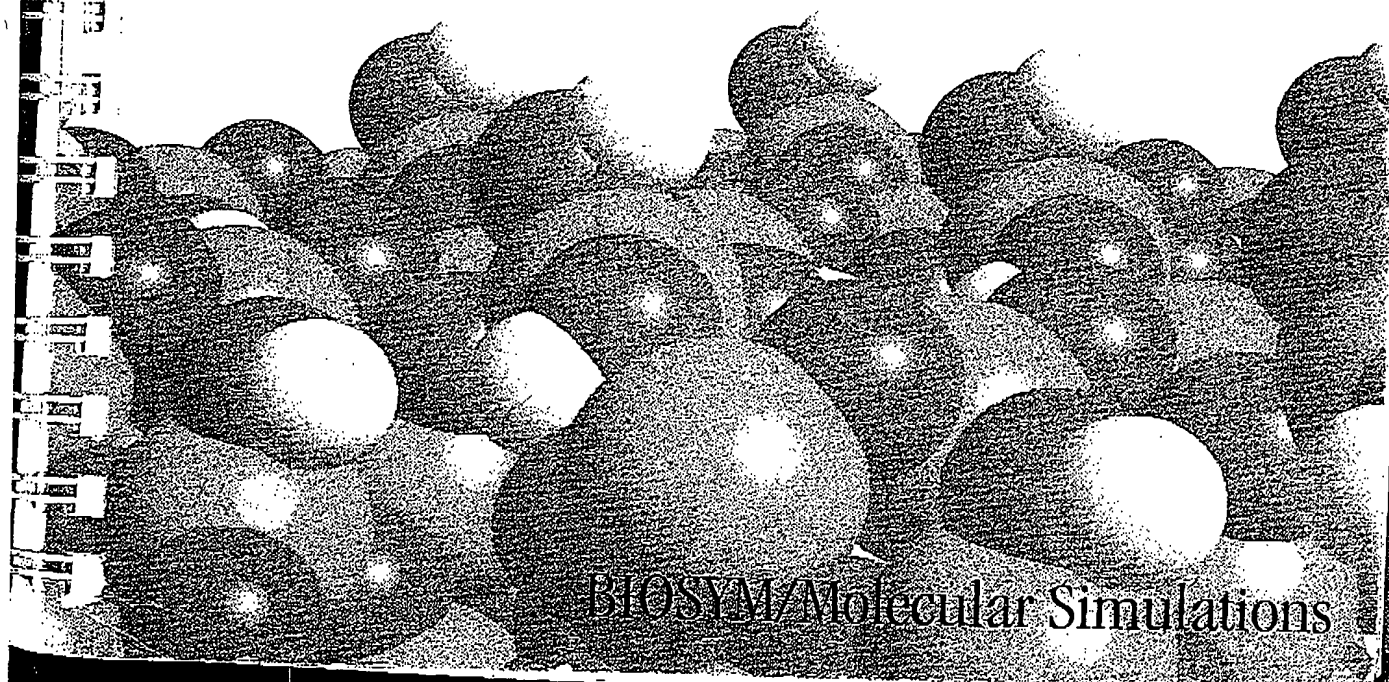
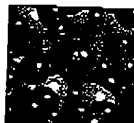
**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

CON #34

# Homology 95.0

## *Protein Structure Prediction*

*User Guide*

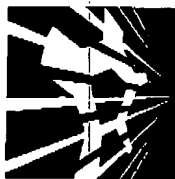


BIOSYM/Molecular Simulations

# *Homology*

## USER GUIDE

*Release 95.0*  
October 1995



Biosym/MSI  
9685 Scranton Road  
San Diego, CA 92121-3752  
619/458-9990 Fax: 619/458-0136

## Copyright\*

This document is copyright © 1995, Biosym/Molecular Simulations. All rights reserved. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means or stored in a database retrieval system without the prior written permission of the publisher.

The software described in this document is furnished under a license and may be used or copied only in accordance with the terms of such license.

### Restricted Rights Legend

Use, duplication, or disclosure by the Government is subject to restrictions as in subparagraph (c)(1)(ii) of the Rights in Technical Data and Computer Software clause at DFARS 252.227-7013.

Biosym/MSI  
9685 Scranton Road  
San Diego, CA 92121-3752  
USA

### Trademark Acknowledgments

*Insight*, *Discover*, Biosym, and the Biosym logo are registered trademarks of Biosym/Molecular Simulations. IRIS and IRIX are trademarks of Silicon Graphics, Inc. UNIX is a registered trademark, licensed exclusively by X/Open Company, Ltd. PostScript is a trademark of Adobe Systems, Inc. The X-Window system is a trademark of the Massachusetts Institute of Technology.

## Permission to Reprint, Acknowledgments, and References

Biosym/MSI usually grants permission to republish or reprint material copyrighted by Biosym/Molecular Simulations, provided that requests are first received in writing and that the required copyright credit line is used. For information published in documentation, the format is "Reprinted with permission from *Document-name*, Month Year, Biosym/MSI, San Diego." For example:

Reprinted with permission from *Homology User Guide*, October 1995, Biosym/MSI, San Diego.

Requests should be submitted to the Biosym/MSI Response Center, either through electronic mail to [rcenter@biosym.com](mailto:rcenter@biosym.com) or in writing to:

Biosym/MSI Response Center  
Biosym/MSI  
9685 Scranton Road  
San Diego, CA 92121-3752

To print photographs or files of computational results (figures and/or data) obtained using Biosym/MSI software, acknowledge the source in the format:

Computational results obtained using software programs from Biosym/MSI of San Diego—dynamics calculations were done with the *Discover*® program, using the CFF91 forcefield, ab initio calculations were done with the *DMol* program, and graphical displays were printed out from the *Insight*® II molecular modeling system.

To reference a Biosym/MSI publication in another publication, no author should be specified and Biosym/MSI should be considered the publisher. For example:

*Homology User Guide*, October 1995. San Diego: Biosym/MSI, 1995.

\*U.S. version of Copyright Page

## Table of Contents

<b>1. Introduction</b>	1-1
What is <i>Homology</i> ?	1-1
Hardware and Installation	1-2
How to Invoke <i>Homology</i>	1-2
Program Environment	1-2
<i>Homology</i> and <i>Insight II</i>	1-2
Saving <i>Homology</i> Information with <i>Insight II</i>	1-2
Command Logging and Restarting	1-3
<i>Homology</i> and <i>Discover</i>	1-3
Operations	1-3
<b>2. Theory</b>	2-1
Background	2-1
<i>Homology</i> Model Building	2-2
Searching Sequence Databases With the FASTA Program	2-2
ktup Value	2-2
Scoring Local Regions	2-3
Joining Regions	2-3
Optimizing the Sequence Matching	2-3
The Determination of Structurally Conserved Regions (SCRs)	2-3
Manual Determination of Structurally Conserved Regions	2-4
Automatic Determination of Structurally Conserved Regions	2-4
Automatic Sequence Alignment Methods	2-8
Needleman and Wunsch Algorithm for Pairwise Alignment	2-9
Biosym's Pairwise Alignment Procedure	2-14
Automatic Multiple Sequence Alignment	2-17
Scoring Matrices	2-25
Multiple Structure Alignment	2-26
Simultaneous Superposition of Structures	2-27
Assignment of Coordinates Within a Conserved Region	2-27
Assignment of Coordinates for Loop or Variable Regions	2-28
Search Loops Command	2-28
Generate Loops Command	2-30
Side Chain Conformational Searches Using Rotamer Libraries	2-33
Refinement of the Model Using Molecular Mechanics	2-34

# A

## References

- Bacon, D.J.; Anderson, W.F. "Multiple sequence alignment," *J. Mol. Biol.*, **191**, 153-161 (1990).
- Barker, W. C.; George, D. G.; Hunt, L. T. "Protein Sequence Database," in *Methods in Enzymology*, **183**, 49, Academic Press: San Diego (1990).
- Barton, G.J. "Protein multiple sequence alignment and flexible pattern matching," *Meth Enzymol*, **183**, 403-428 (1990).
- Bencdetti, E.; Morelli, G.; Nemethy, G.; Scheraga H. A. "Statistical and Energetic Analysis of Side-chain Conformations in Oligopeptides," *Int. J. Peptide Protein Res.* **22**, 1-15 (1983).
- Berger, M.P.; Munson, P.J. "A novel randomized iterative strategy for aligning multiple protein sequences," *Comput Appl Biosci*, **7**, 479-484 (1991).
- Blundell, T. L.; Sibanda, B. L.; Sternberg, M. J. E.; Thornton, J. M. "Knowledge-based prediction of protein structures and the design of novel molecules," *Nature*, **326**, 347 (1987).
- Blundell, T.L.; Carney, D.; Gardner, S.; Hayes, R.; Howlin, B.; Hubbard, T.; Overington, J.; Singh, D. A.; Sibanda, B. L.; Sutcliffe, M. "Knowledge-based protein modelling and design," *Eur. J. Biochem.*, **172**, 513 (1988).
- Browne, W.J.; North, A.C.T.; Phillips, D. C.; Brew, K.; Vanaman, T. C.; Hill, R. L. "A Possible Three-dimensional Structure of Bovine  $\alpha$ -Lactalbumin based on that of Hen's Egg-White Lysozyme," *J. Mol. Biol.*, **42**, 65 (1969).
- Burks, C. et al. "GenBank: Current Status and Future Directions," in *Methods in Enzymology*, **183**, 3-4, Academic Press: San Diego (1990).
- Chothia, C. "Hydrophobic Bonding and Accessible Surface Area in Proteins," *Nature*, **248**, 338-339 (1974).
- Chothia, C. "The Nature of the Accessible and Buried Surfaces in Proteins," *J. Mol. Biol.*, **105**, 1-14 (1976).
- Connolly, M.L. "Solvent-Accessible Surfaces of Proteins and Nucleic Acids," *Science*, **221**, 709-713 (1983).
- Dayhoff, M. O.; Barker, W. C.; Hunt, L. T. "Establishing Homologies in Protein Sequences," *Methods in Enzymology*, **91**, 524 (1983).

## References

- Dayhoff, M.O.; Schwartz, R.M.; Orcutt, B.C. "A model of evolutionary change in proteins." In: *Atlas of protein sequence and structure*, Dayhoff M.O., Ed., Washington: Natl. Biomed. Res. Found., Vol. 5, Suppl. 3, 345-352 (1978).
- Depiereux, E.; Feytmans, E. "Simultaneous and multivariate alignment of protein sequences: correspondence between physicochemical profiles and structurally conserved regions (SCR)," *Protein Engng*, 4, 603-613 (1991).
- Dill, K. "Dominant Forces in Protein Folding," *Biochem.*, 29, 7133-7155 (1990).
- Eisenberg, D.; Weiss, R.M.; Terwilliger, T.C.; Wilcox, W. "Hydrophobic Moments and Protein Structure," *Faraday Symp. Chem. Soc.*, 17, 109-120 (1982).
- Eisenberg, D.; McLachlan, A.D. "Solvation Energy in Protein Folding and Binding," *Nature*, 319, 199-203 (1986).
- EMBL Data Library; European Molecular Biology Laboratory, Postfach 10.2209, 6900, Heidelberg, Germany
- Engelman, D. M.; Steitz, T. A. "The spontaneous insertion of proteins into and across membranes: the helical hairpin hypothesis," *Cell*, 23, 411 (1981).
- Engelman, D. M.; Steitz, T. A.; Goldman, A. "Identifying Nonpolar Transbilayer Helices in Amino Acid Sequences of Membrane Proteins," *Ann. Rev. Biophys. Chem.*, 15, 321 (1986).
- Garnier, J.; Robson, B. in *Prediction of Protein Structure and the Principles of Protein Conformation*, Fasman, G., Ed., Plenum: New York, Ch. 10, 417-465 (1989)
- GenBank database; IntelliGenetics, Inc., 700 El Camino Real East, Mountain View, CA 94040
- Greer, J. "Model for haptoglobin heavy chain based upon structural homology," *Proc. Nat. Acad. Sci. U.S.A.*, 77, 3393 (1980).
- Greer, J. "Comparative Model-building of the Mammalian Serine Proteases," *J. Mol. Biol.*, 153, 1027 (1981).
- Greer, J. "Model Structure for the Inflammatory Protein C5a," *Science*, 228, 1055 (1985).
- Hopp, T.P.; Woods, K. "Prediction of protein antigenic determinants from amino acid sequences," *Proc. Natl. Acad. Sci. USA*, 78, 3824-3828 (1981).
- Jackson, R.M.; Sternberg, M.J. "Protein Surface Area Defined," *Nature*, 366, 638 (1993).
- Janin, J. "Surface and inside volumes in globular proteins," *Nature*, 277, 491-492 (1979).
- Johnson, M.S.; Doolittle, R.F. "A method for the simultaneous alignment of three or more amino acid sequence," *J. Mol. Evol.*, 23, 267-278 (1986).



## References

- Kabsch, W.; Sander, C. "Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features," *Biopolymers*, **22**, 2577-2637 (1983).
- Kahn, P.; Cameron, G. "EMBL Data Library," in *Methods in Enzymology*, **183**, 26, 31, Academic Press: San Diego (1990).
- Karlin, S.; Altschul, S.F. "Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes," *Proc. Natl. Acad. Sci. USA*, **87**, 2264-2268 (1990).
- Karlin, S.; Dembo, A.; Kawabata, T. "Statistical composition of high-scoring segments from molecular sequences," *Ann. Stat.*, **18**, 571-581 (1990).
- Karlin, S.; Brendel, V. "Chance and statistical significance in protein and DNA sequence analysis," *Science*, **257**, 39-49 (1992).
- Kyte, J.; Doolittle, R. F. "A Simple Method for Displaying the Hydrophobic Character of a Protein," *J. Mol. Biol.*, **157**, 105-132 (1982).
- Lee, B.; Richards, F.M. "The Interpretation of Protein Structures: Estimation of Static Accessibility," *J. Mol. Biol.*, **55**, 379-400 (1971).
- Lipman, D. J.; Pearson, W. R. "Rapid and Sensitive Protein Similarity Searches," *Science*, **227**, 1435-1441 (1985).
- Lipman, D.J.; Altschul, S.F.; Kececioglu, J.D. "A tool for multiple sequence alignment," *Proc. Natl. Acad. Sci. USA*, **86**, 4412-4415 (1989).
- Mas, M.T.; Smith, K.C.; Yarnush, D.L.; Aisaka, K.; Fine, R.M. "Modeling the anti-CEA antibody combining site by homology and conformational search," *Proteins: Struct., Func., and Genet.*, **14**, 483-498 (1992).
- McGregor, M. J.; Islam, S. A.; Sternberg, M. J. "Analysis of the Relationship between Side-chain Conformation and Secondary Structure in Globular Proteins," *J. Mol. Biol.*, **198**, 195-210 (1987).
- Murata, M.; Richardson, J.S.; Sussman, J.L. "Simultaneous comparison of three protein sequences," *Proc. Natl. Acad. Sci. USA*, **82**, 3073-3077 (1985).
- Needleman, S. B.; Wunsch, C. D. "A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins," *J. Mol. Biol.*, **48**, 443-453 (1970).
- Novotny, J.; Brucoleri, R.; Karplus, M. "An Analysis of Incorrectly Folded Protein Models," *J. Mol. Biol.*, **177**, 787-818 (1984).
- Ooi, T.; Oobatake, M.; Nemethy, G.; Scheraga, H. "Accessible Surface Areas as a Measure of the Thermodynamic Parameters of Hydration of Peptides," *Proc. Natl. Acad. Sci.*, **84**, 3086-3090 (1987).